

FINAL PROJECT OVERVIEW

Stat 120 | Fall 2025

Prof Amanda Luby

What is the project? The final project will be a short paper + mini-poster presentation on any reasonable topic involving data analysis or statistics. Posters will be displayed in the style of a professional conference, scheduled for the last day of class. Projects are to be done in groups of three or four students (groups of 2 or 5 may be allowed). The paper component of the project is a 3-page overview of your analysis and results, and the poster presentation includes the physical poster along with a 2-minute “elevator pitch” of your project. A proposal (worth part of your project grade) is due before the midterm break.

The goal is not to do an exhaustive data analysis i.e., do not calculate every statistic and procedure you have learned for every variable, but rather let me know that you are proficient at asking meaningful questions and answering them with results of data analysis, that you are proficient in using R, and that you are proficient at interpreting and presenting the results. Focus on methods that help you begin to answer your research questions, and make sure your data is appropriate for the methods that we’ve learned about in class. Also, critique your own methods and provide suggestions for improving your analysis. Issues pertaining to the reliability and validity of your data, and appropriateness of any statistical analysis should be discussed.

What are possible types of projects? A project may be one of the following:

1. An analysis of an existing dataset. You might apply some technique we have learned in the course to a dataset that has already been collected. I recommend choosing this option and I anticipate most projects to be in this category. For some ideas:
 - FiveThirtyEight: data.fivethirtyeight.com
 - DataIsPlural: data-is-plural.com
 - HelpMeViz: policyviz.com/helpmeviz/
 - TidyTuesday: github.com/rfordatascience/tidytuesday
 - Urban Institute’s Data Catalog: datacatalog.urban.org
 - Data from studies published in *Nature*: nature.com/sdata/policies/repositories
2. A research report on a topic that interests you – perhaps a topic we have touched on in class, or something you learned about in another class or through your own reading or experience. Such a project might involve reading articles in journals, newspapers, magazines, and books; summarizing and discussing the statistical issues involved; and drawing conclusions. Examples of possible topics include algorithmic bias, the use of DNA in court cases, the use of statistics in standardized testing, the probabilities involved in gambling, the use of election polls in the media, color theory for statistical graphics, the use of regression and forecasting for the stock market, etc. You must engage with the content of the course, and so please be in touch with me early to make sure your plan is appropriate.
3. Collecting your own data and analyzing it. If you choose this option, your research question should be clearly motivated and you must justify the need to collect new data. These types of projects work best if your group is interested in a specific research question that can be answered through an observational study or experiment. (Examples: observing parking use on campus; how style of frisbee throw impacts distance/accuracy; surveys of plants in the arboretum, etc.) As we’ve seen, it

is incredibly difficult to collect a random sample using online surveys, and so I will only allow those types of projects if discussed with me in advance, and the sampling used will be part of your grade.

If you have another project idea that does not fall into one of these categories, feel free to talk to me. I encourage interesting and innovative ideas!

What kind of data should we use? As a general minimum, your dataset should include at least five variables, contain at least 30 observations, and include a mix of quantitative and categorical variables. You should study at least two questions.

There's *lots* of data available freely online. In general, you're welcome to use any dataset. You are responsible for knowing where your data came from (is it real? was it collected ethically? what is the population and what is the sample?)

Please keep in mind that many datasets have complex structures that require methods that are more advanced than we've learned in this class. Your dataset should represent a simple random sample of a population. Some examples of other data that generally won't work well in this setting are:

- Time series data (measurements of a single unit over time)
- Spatial data
- Data representing the whole population of interest

What should the paper include? We will be following the guidelines for the [Undergraduate Statistics Class Project Competition](#). The 3-page paper should include the following components:

(1) *A statement of the research question or purpose.* What problems or questions did you set out to investigate? What were the key issues raised?

(2) *The background and significance of the project.* What sources or background readings did you consult? What information did you use in developing your ideas from the conceptual stage to the finished project? What questions does your project answer, and why is it important?

(3) *Methods.* What statistical techniques did you use — for instance, scatterplots, correlation, hypothesis tests, confidence intervals, regression, ANOVA? Did you gather information using a survey, experiment, or other method?

(4) *Results and conclusions,* the summary and presentation of your data analyses. What did you find out? This might include tables, graphs, or verbal summaries. What did you learn about the problem or question you set out to investigate?

(5) *Discussion and critique.* What did you learn about the process of carrying out your project? What are the limitations of your project? How reasonable are the assumptions you made? What future work/studies would you recommend?

Not all points will be relevant to all projects. For example, research reports may not have any “results” as such. For some examples, check out the [previous winners](#) of the USCLAP.

What should the poster include? Your poster should be a summary of your main findings. Your poster should be 11”x17”. You may use any template/format for the poster (I’ll provide a google slides template later in the term). You may print it on your own, or I will print it for you if you send me a PDF 24 hours in advance of the poster session. I recommend printing early to avoid last-minute technology failures.

Your group should also prepare a verbal “elevator pitch” of your presentation and be prepared to answer questions about your project.

How will the project be graded? All group members receive the same grade (except in extreme circumstances). The project is worth 25% of your final course grade (10% for the poster and 15% for the paper). A more detailed rubric will be provided later in the course, but you can expect to be evaluated on:

- *Statistical Content:* Does the project incorporate concepts and methods from this course? Are they used appropriately and correctly?
- *Subject matter content:* Did you engage with the subject of the data/project? What did you learn?
- *Presentation and formatting:* Is the poster/paper attractively presented and well organized? Are all graphs, equations, and tables readable?
- *Meeting due dates*
- *Peer Review of Posters*

How do I find a project group? Fill out the google form as soon as possible. I will assign groups based on the results as soon as possible.

What are the deadlines?

Week 4 Friday @ start of class: Fill out google form for forming groups. I will assign groups by Monday of Week 5.

Week 5 Friday @ 5pm: Project Proposal due on gradescope. Please submit one proposal as a group, but make sure all group members are included on the submission. The format for the proposal is below.

Last Day of Class: Poster Session! If you’d like me to print your poster, please email it to me 24 hours in advance.

Last day of finals @ noon: Final papers due

A final word: I encourage innovation, creativity, and thoughtfulness in your projects. I want you to have fun with this assignment, explore something you’re interested in, and create a piece of work that you’re proud of. If you have an idea that doesn’t neatly fit into these prompts, please talk to me! If you have a topic you’re interested in but no idea what a project might look like, please talk to me!

Project Proposal Template: You should answer the following questions in a PDF document submitted to gradescope. Only submit 1 project proposal per project; project members should be listed on the document and included in the file submission.

1. Names of group members:
2. General Topic:
3. Depending on the type of project chosen:
 - If analyzing an existing dataset, provide a link or attach a file with the dataset. State what each *case* is and descriptions of at least 3 *variables* you are interested in investigating. You should include the R code you use to load your dataset.
 - If doing a research project, include at least 3 sources (articles, books, etc.) that you will use and a short description of how you plan to organize your paper/presentation.
 - If collecting data, provide a detailed plan for how you will collect AND analyze the data, including a detailed timeline.

Homework “BYOD” problems: On most homeworks from now until the end of the term, one R problem will have an alternative “BYOD” (bring your own data) option. You may replace the analyses in these questions with one from your final project data (if appropriate). This is not required, but can be a great way to make progress on your final project analyses as part of your homework.

Extra Credit: There will be some extra credit on the final paper portion of the project if your group submits to the USCLAP competition. To be eligible for the extra credit, your group must submit a rough draft of your paper by midnight on Monday Week 10. I will get comments back to you by Friday, and your final paper should address those comments and make any requested changes.